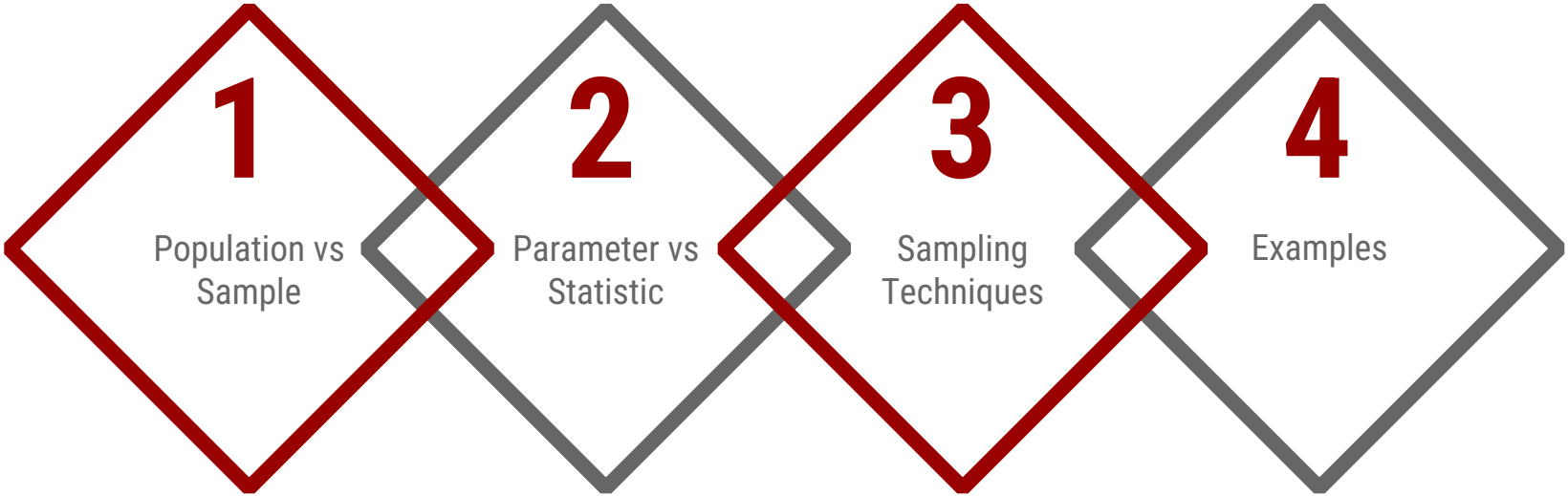
# 8.1 Collecting Data

# Goals for the Day

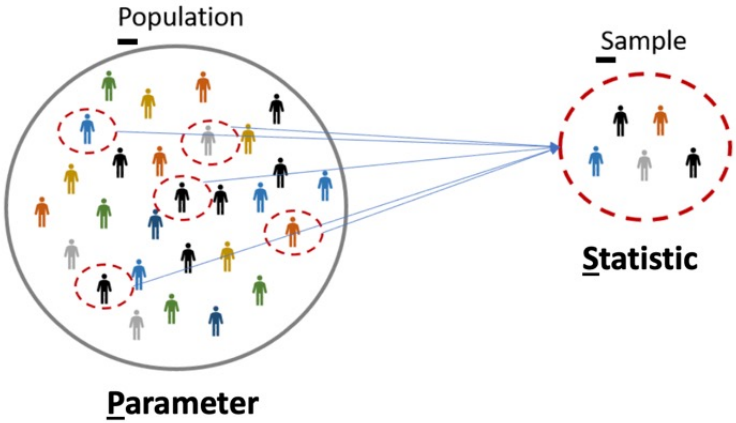


# 1

# Population vs Sample

# What's the difference?

- **Population** – the particular group of interest in a study
  - The set of all individuals/objects of interest
- **Sample** – a subset of the individuals/objects from the population of interest



# What's the difference?



Example

Let's say we want to know if Indiana is a cat or dog state.

What is the population? Every person in ALL of Indiana

What is the sample? Everybody in Muncie ONLY

# 2

# Parameter vs Statistic

# What's the difference?

- (Population) Parameter – a fixed numerical value that describes the population
  - EX: Percentage of people in all of Indiana who prefer cats
  - Would need to take a **census** (ask everyone in the population) to know this value (or estimate it)
- (Sample) Statistic – a numerical value that describes the sample that can vary from sample to sample
  - EX: Percentage of people in Muncie who prefer cats (will be different than for Indy)

# 3

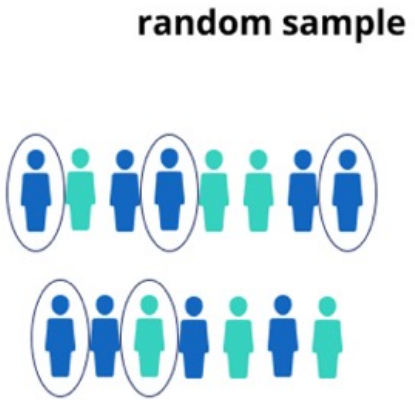
# Sampling Techniques

# Sampling

- Need to find a way to pick who/what is going to be included in the sample
- Goal: A **representative sample** – a sample that has the same relevant characteristics as the population AND does not favor one group of the population over another.
- Many different methods for sampling

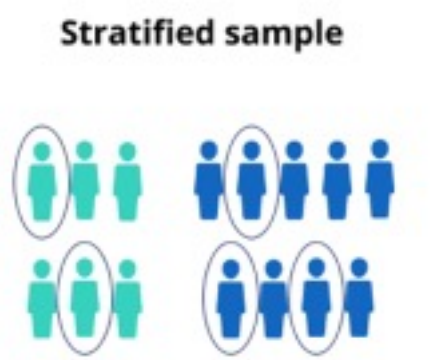
# Sampling Strategies

- **Random Sample** – every member of the population has an equal chance of being selected
  - This is generally desirable but can be difficult to achieve.



# Sampling Strategies

- Stratified Random Sample – dividing the population into homogeneous (similar characteristics) groups
  1. Stratify the population - divide the population into similar groups (e.g., based on age or gender)
  2. Take random sample from each group (strata)
  3. Combine the groups from each strata to form your sample

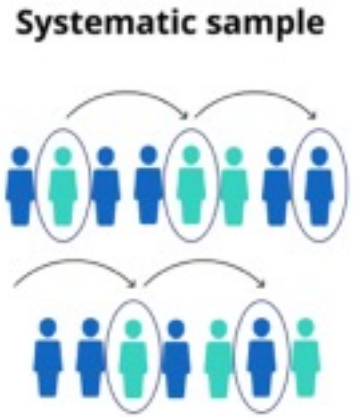


# Sampling Strategies

- Cluster Sample – dividing the population into mini-populations. This gives us an unbiased sample and is often a more practical / affordable method.
  1. Split the population into representative groups called clusters (should resemble overall population)
  2. Use random sampling to select several whole clusters
  3. Perform a census of each selected (collect data from every member).

# Sampling Strategies

- Systematic Sample – selecting every $n^{th}$ member of the population



# Sampling Strategies

- Convenience Sample – include individuals who are convenient to sample (for the researcher); AVOID!
  - This group may not be representative of population
  - Frequently leads to biased results

# 4

# Examples

# Example #1

Describe how you could obtain a sample to answer the question below using each of the following types of sampling methods.

I want to determine the proportion of MATH 125 students that has a Mac laptop.

**Random** – Randomly ask 10 students from class

**Cluster** – Randomly sample 3 tables; ask everyone in that (take a census of the) table

**Convenience** – Ask the 5 students closest to me

**Stratified** – Randomly sample 5 students from the list of Freshman and Sophomores respectively

**Systematic** – choose every $4^{th}$ student off the roster

# Example #2

Describe how you could obtain a sample to answer the question below using each of the following types of sampling methods.

You are tasked with conducting a survey to answer the question, "What is the favorite subjects of students who attend East High School?"

**Random** – Get a list of all student names and randomly select 10 names

(could number each name and randomly generate 10 numbers)

**Stratified** –Divide students by grade level OR by social group  band, football, etc.) and then randomly sample within each group

(Students are the same WITHIN each group, but different ACROSS groups)

**Cluster** – Randomly sample 5 classrooms OR buses, and then ask every student in each.

(Should be a mix of students in each cluster, all clusters → same)

**Systematic** – Ask every 5th student that arrives in the morning.

Get a list of all students names and select every 10th name.

**Convenience** – Walk through hallway and ask first 20 students you pass.

(Bad because maybe I am by the chem lab and only ask chem students; their opinions might not match the overall student body's opinion